Flexible Analysis of Electronic Medical Record Data with Composite Mixture Models

Michael Mayhew[1], Brenden Petersen[1,] Ana Paula Sales[1], John Greene[2], Vincent Liu[2], Todd Wasson[3]
[1]Computational Engineering Division, Lawrence Livermore National Laboratory
[2]Division of Research, Kaiser Permanente Northern California
[3]LinkedIn, Inc.

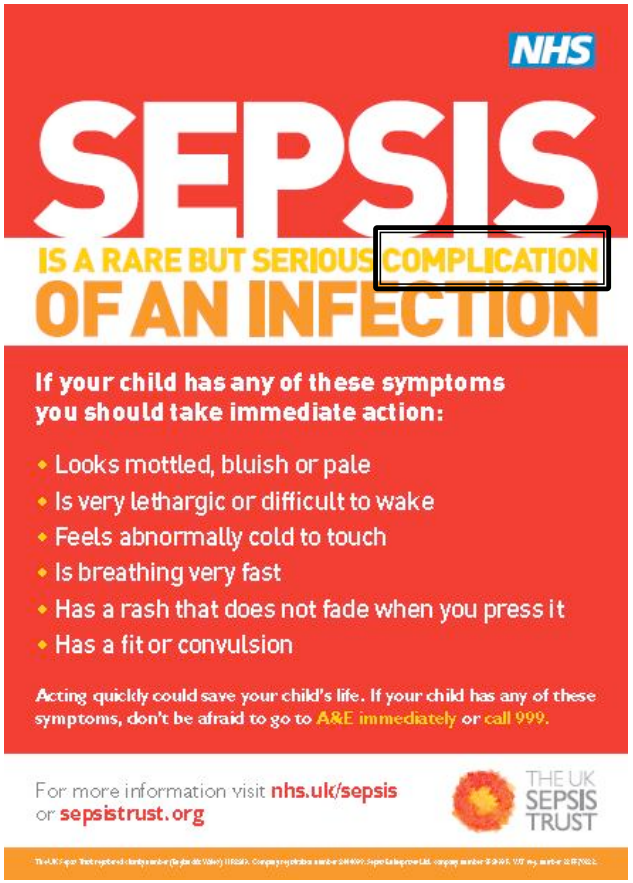
CASIS
May 17, 2017

LLNL-PRES-731012
This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. Lawrence Livermore National Security, LLC

Lawrence Livermore National Laboratory

Towards Mortality Risk Stratification in Sepsis

- Sepis is *costly*
  - As of 2011, sepsis affects nearly 1 million Americans in hospitals (28-50% mortality) and costs nearly $20 billion every year
- Sepsis is *heterogeneous* in its presentation
  - Elderly, young, immuno-compromised and those with pre-exsiting conditions are particularly at risk
- Sepsis is *difficult* to recognize and treat
  - Fever, chills, rapid breathing and heart rate, disorientation & confusion
  - Mortality increases by 7.6% with every hour anti-microbial administration is delayed after onset of hypotension (Kumar et al., 2006)

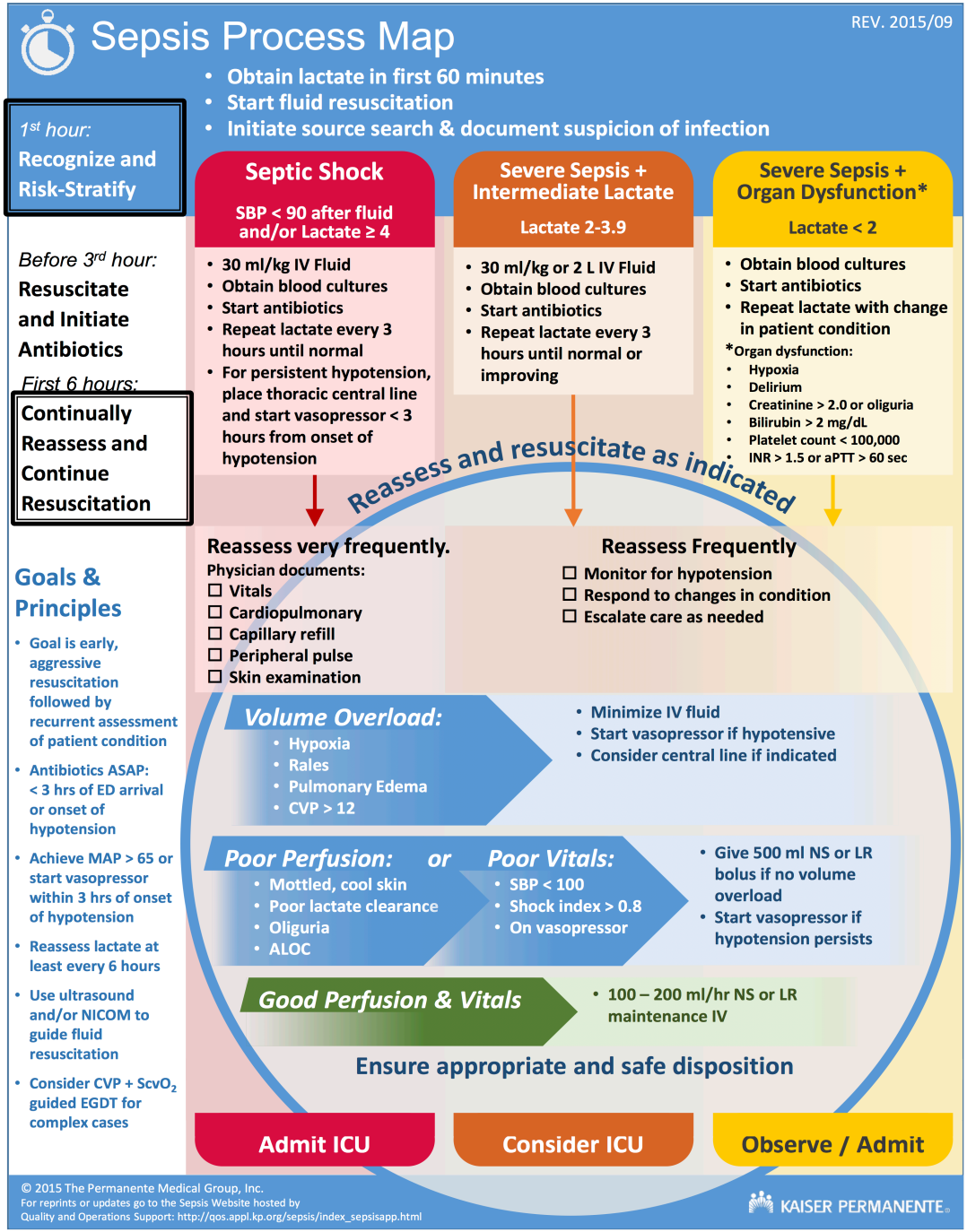


https://www.nigms.nih.gov/education/pages/factsheet_sepsis.aspx

http://sepsistrust.org

Sepsis Detection and Treatment at Kaiser Permanente Northern California

- Clear and present need to:
  - Stratify patients accurately using all relevant observations
  - Detect physiological changes associated with adverse outcomes as early as possible
  - Identify *clinically actionable* signatures of physiological deterioration of patient



Composite Mixture Models: Flexible Analysis of Multi-Typed Data from Heterogeneous Populations

$$\Pr(x|\Theta) = \sum_{k=1}^{K} \pi_k \prod_{i=1}^{|x|} \Pr_i(x_i|\theta_{k,i})$$

Proportion of population belonging to component *k*

Parameters specific to component *k* and dimension *i*

- We assume that:
  1. Population is heterogeneous (can be divided into subgroups or components)
  2. Each observation dimension can be modeled with an appropriate univariate, exponential family distribution ($\Pr_i$)
  3. Observation dimensions are independent of one another given mixture component

- Complex dependencies can be recovered by averaging over a sufficient number of mixture components

Sales et al., 2013; Wasson & Sales, 2014

Description of KPNC EMR Analysis Cohort

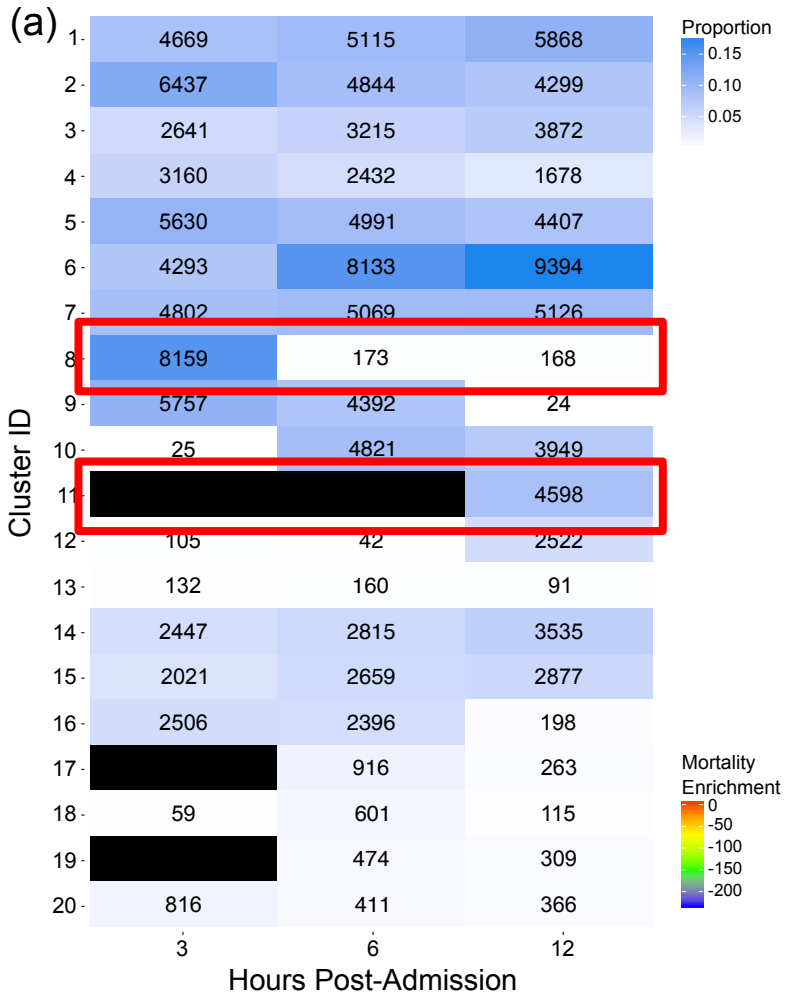
- EMR observations:
  - Are both static (e.g. gender) and dynamic (e.g. patient vital signs)
  - Are both discrete (e.g. diagnosis codes) and continuous (e.g. acute disease burden)
  - Tend to have missing entries

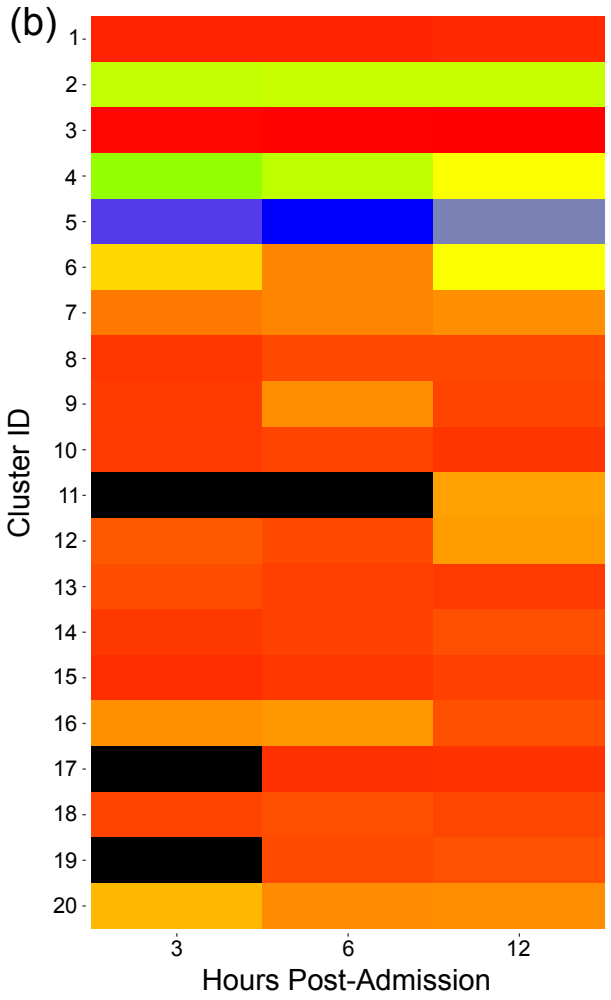
- 53,659 emergency department hospitalizations (~46k patients)
  - Minimum of 12-hour stay
  - At least 3 vitals obs. at 3 hours of stay
  - Patient had suspected or confirmed infection
  - Mortality rate: 6.4%, Mean age: 67yo, 51% female

- Admission/Demographic features:
  - age, sex
  - LAPS2, COPS2 – measures of acute & chronic illness, respectively
  - Kaiser-specific variables (Facility code, Membership indicator, Transport from non-KP facility indicator)

- Vitals features:
  - Median, max, min, and standard deviation at 3, 6 and 12 hours post-admission

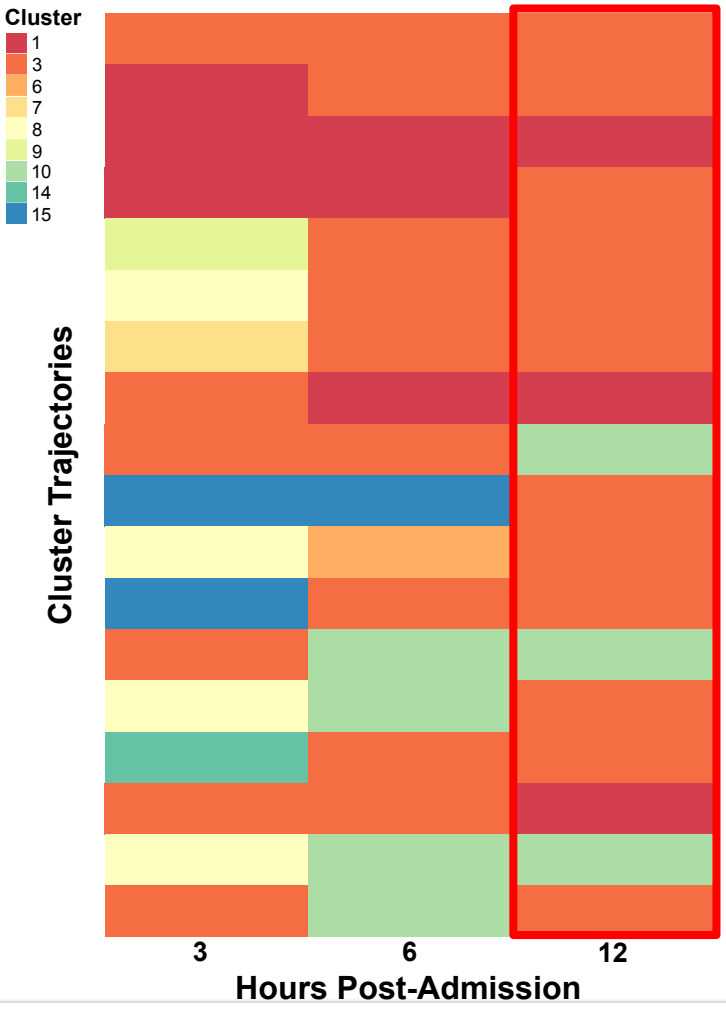
Inferred Physiological Space is Complex





- Patients associated with different clusters over hospitalization
- Some clusters appear transient
- Clusters show different enrichment for mortality events

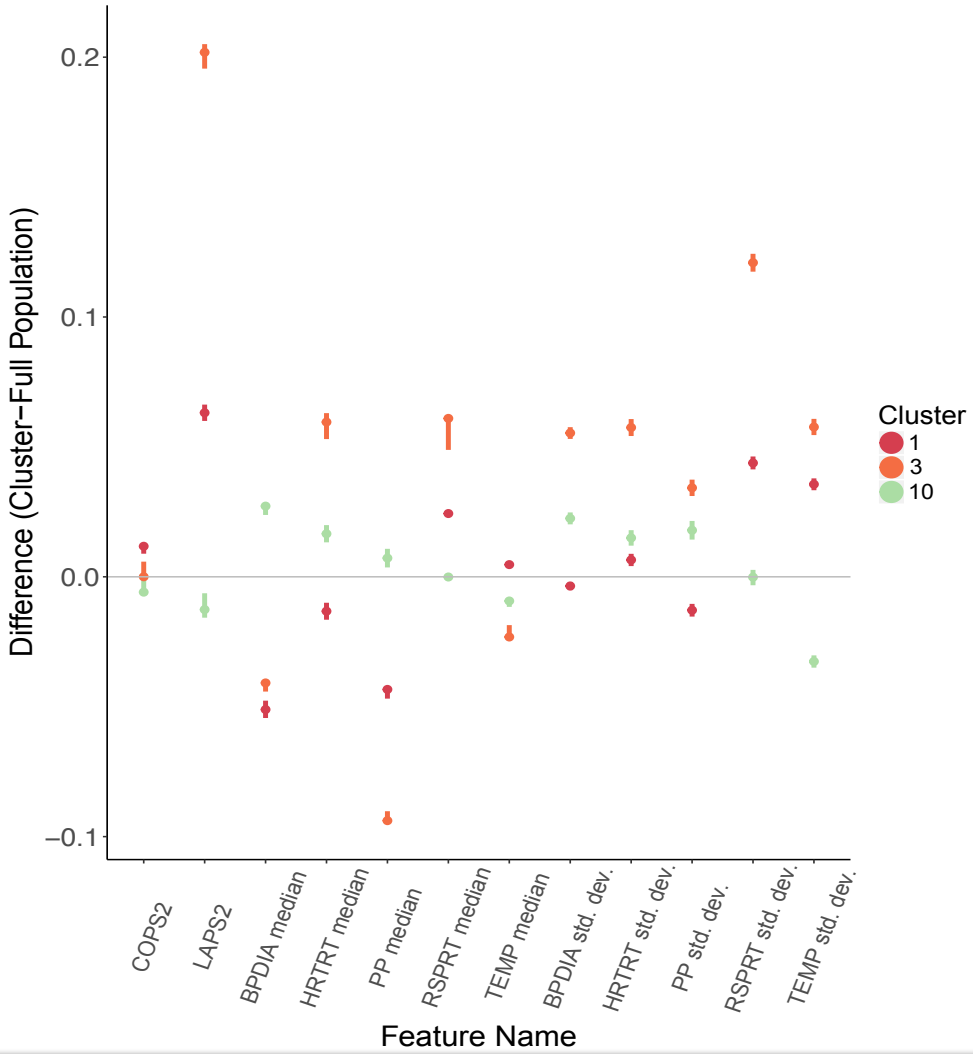
CMMs Reveal Temporal Patterns that Aid Risk Stratification of Septic Patients



-log p-value	no. of deaths	total	mortality rate
876.18	717	1848	0.39
65.03	66	210	0.31
54.13	229	1782	0.13
45.90	45	142	0.32
45.67	47	155	0.30
32.36	46	202	0.23
24.17	21	64	0.33
23.11	31	135	0.23
21.21	13	29	0.45
19.58	19	66	0.29
18.45	19	70	0.27
16.71	20	84	0.24
16.55	8	15	0.53
15.77	19	81	0.23
14.52	19	87	0.22
11.97	17	85	0.20
11.67	122	1289	0.09
11.35	5	10	0.50

- Cluster trajectories consist of joint cluster assignments at 3, 6 and 12 hours post-admission of each episode
- Shown are top 18 trajectories associated with mortality events

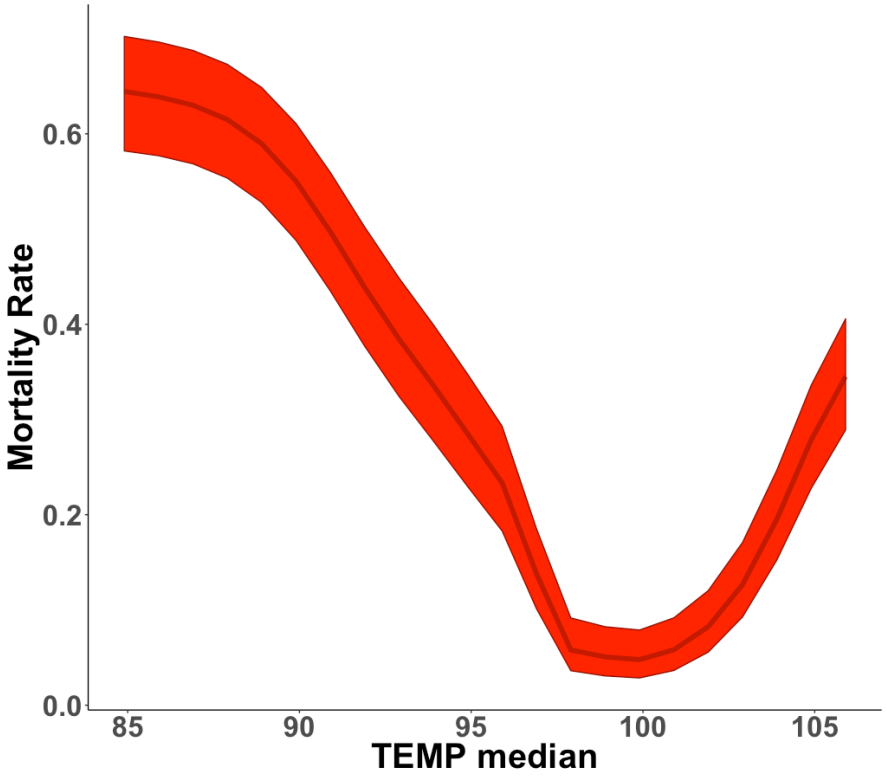
CMM-Based Cluster Analysis Identifies Physiologically Distinct Sub-Populations



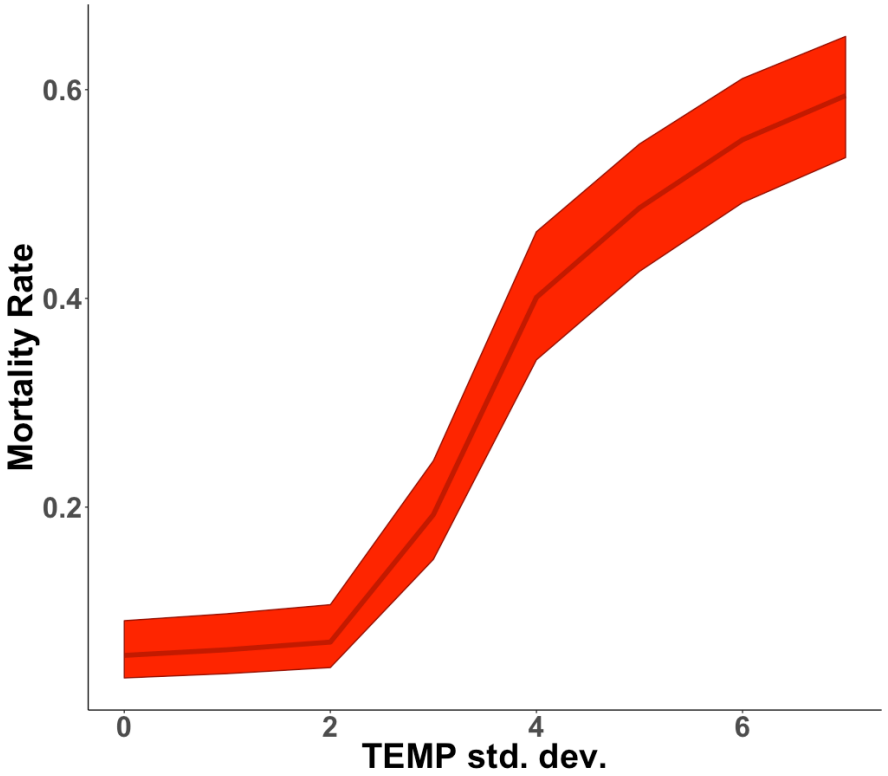
- LAPS2 and COPS2 are KPNC measures of *acute* and *chronic* disease burden, respectively
- Shown are estimated differences between physiological features from mortality-enriched CMM clusters 1, 3, and 10 at 12h post-admission

CMMs Enable Visualization of Physiological Trends Associated with Elevated Mortality Risk

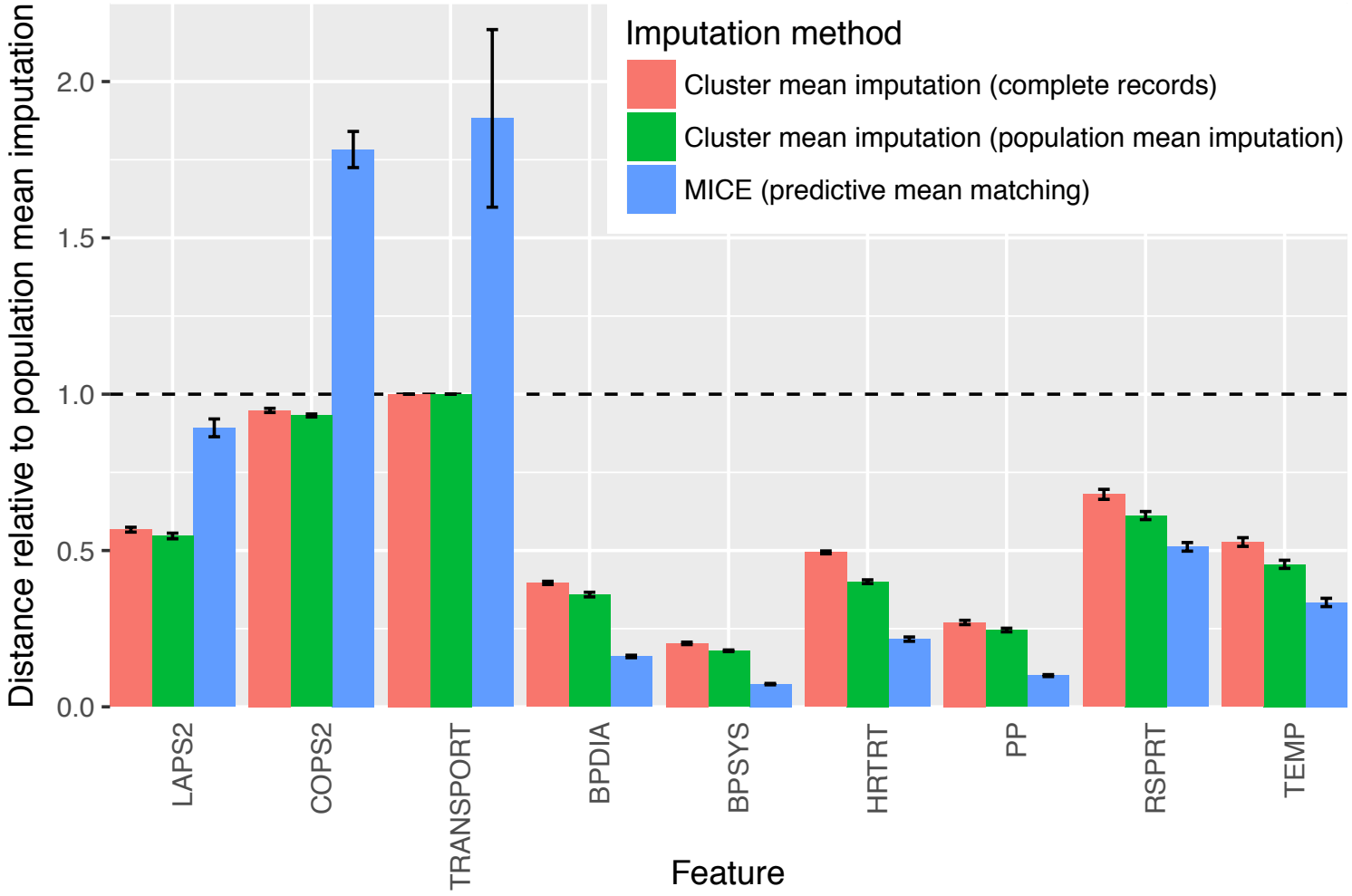
Median Body Temperature (12h Post-Admission)



Body Temperature Std. Dev. (12h Post-Admission)



CMMs Can Enhance Missing Data Imputation Performance



Acknowledgements

LLNL

- Ana Paula Sales
- Todd Wasson (LinkedIn, formerly of CED/ENG)
- Priyadip Ray
- Brenden Petersen

KPNC

- Dr. Vincent Liu
- John D. Greene
- Jennifer Baker


Lawrence Livermore
National Laboratory